

MUESTRA EN INVESTIGACION CUANTITATIVA

THE SAMPLE IN QUANTITATIVE RESEARCH

Juan Carlos Argibay*

Resumen

En el siguiente artículo se abordan cuestiones relacionadas con el tema de la “muestra” en investigaciones cuantitativas. Se comentan los requisitos necesarios para que una muestra pueda ser considerada representativa. Se plantean las diferencias entre muestreos probabilísticos y no probabilísticos, se analizan las cuestiones de orden práctico que hacen que en investigación muchas veces se utilicen muestras incidentales y cómo esto puede afectar -o no- la validez externa de los distintos tipos de métodos de investigación cuantitativos y qué otros recursos se pueden utilizar para minimizar ese problema. Se discute también cómo el tamaño de la muestra excede el tema de la representatividad de la misma y se plantea cómo está vinculado con cuestiones metodológico-estadísticas. Se explica por qué la cantidad de sujetos que componen la muestra es un tema importante en distintas investigaciones y cómo el tamaño de la muestra se vincula con la significación estadística.

Palabras clave: muestra, muestra representativa, técnicas de muestreo, tamaño de muestra, diseño, significación estadística.

Summary

The following article deals with questions related to the subject of the sample in quantitative researches. It shows the requirements to consider a sample representative. Besides, the difference between probabilistic samples and not probabilistic ones is explained. Moreover, the practical matters which explain the often used incidental samples in research as well as how they can affect the external validity of the different sorts of quantitative research methods (or not) and which other resources can be used to minimize that problem are considered. The article also discusses how the size of the sample exceeds the subject of its representativeness and explains how this is related to methodological and statistical questions, why the number of people composing the sample is an important issue in different researches and how the size of the sample is connected with the statistical significance.

Key words: sample, representative sample, sample techniques, size of sample, design, statistical significance.

* Licenciado en Psicología. Profesor Adjunto en Metodología de la Investigación Psicológica de la Licenciatura en Psicología UBA. Profesor Adjunto en Metodología III, del Doctorado en Psicología y Profesor invitado en el Seminario de Metodología de la Maestría en Problemas y Patologías del Desvalimiento UCES. Profesor Titular de Estadística Aplicada a la Educación y Profesor Titular de Metodología de la Investigación en la Licenciatura y en el Profesorado en Ciencias de la Educación UCALP. Dirección: Lugones 4777 (1430), Ciudad de Buenos Aires, Argentina. E-mail: jcargibay@hotmail.com

Introducción

El objetivo de este artículo es plantear algunas cuestiones que tienen que ver con la “muestra” de sujetos que participan en una investigación. Específicamente me referiré a aquellas cuestiones que pueden surgir en investigaciones cuantitativas. No es objetivo del artículo abordar las problemáticas que se pueden dar en las investigaciones cualitativas. Tampoco nos referiremos a las investigaciones cuantitativas de sujeto único, ya que a ellas no se aplican las cuestiones de muestreo tal cual las enfocaremos. Los diseños de sujeto único, son una forma especial de diseño en los que se trabaja con un solo sujeto, o si se hace con más de uno, pero siempre en número limitado, como en los diseños de línea de base múltiple a través de individuos, el enfoque sigue estando en el la unidad.

Nos referiremos a varios temas, uno tiene que ver con las técnicas de muestreo empleadas, y de qué manera estas nos aseguran la calidad de la muestra; otros están vinculados con el número de sujetos que participan de la investigación. Esto último en lo que hace a que la muestra pueda ser considerada representativa; pero también en lo que afecta a cuestiones metodológicas y estadísticas, como ser por ejemplo la viabilidad de determinados diseños de investigación. El tamaño de la muestra está directamente vinculado con el tema de la significación estadística y con que podamos corroborar o no las hipótesis que planteamos en nuestras investigaciones. Abordaremos cómo, en investigaciones experimentales, la adecuada formación al azar de los grupos experimentales y de control, para garantizar la equivalencia inicial de los mismos, también guarda relación, con el tamaño de la muestra.

Técnicas de muestreo. Representatividad de la muestra. Validez externa

En cualquier investigación que realicemos, debemos tomar una muestra de sujetos que serán quienes participen de la misma. En general se habla mucho en estos casos acerca de que la muestra que se tome debe ser representativa. Esto es así, ya que este es uno de los aspectos que serían necesarios para asegurar la validez externa de la investigación. “*La validez externa se halla asociada a la generalización y representatividad de los logros de la investigación*”, Arnau Gras, J. (1982) (pág. 351). Si investigamos, no es únicamente para enterarnos qué pasa con esa muestra de sujetos en particular, sino que el objetivo es poder extender esos resultados a otros sujetos y situaciones, de ahí que resulte de fundamental importancia el tema de la validez externa y uno de sus aspectos que es, si la muestra que hemos utilizado es suficientemente representativa de la población de referencia para poder extender los resultados obtenidos en aquella a esta. Cabe aclarar que dentro de la validez externa también tenemos lo que se da en llamar validez ecológica, o sea, “*la posibilidad de generalizar a una situación natural los resultados obtenidos en otra artificial*”, Pereda Marín, S. (1987) (pág. 261).

Hay que tener en cuenta entonces que el hecho de que la muestra de sujetos empleada en la investigación sea representativa, no asegura en un sentido amplio la posibilidad de generalización de la misma, solo en lo que respecta a la posibilidad de generalizar los resultados a determinada población. Podría ocurrir por ejemplo, que la muestra sea representativa, pero que tenga muy poca validez ecológica, de manera que trasladados los resultados a situaciones naturales, no sean reproducibles. Dicha investigación tendería, a pesar de la muestra representativa, problemas de validez externa. ¿Cuál sería la

importancia de saber que los sujetos que participaron de la investigación son equivalentes a los de la población a la cual se pretende generalizar, si de todos modos, los resultados obtenidos no serían trasladables a situaciones externas al contexto de la investigación? Del mismo modo, si la investigación tuviera muy buena validez ecológica, de manera que sus resultados fueran perfectamente reproducibles en situaciones naturales, pero, únicamente para un grupo de personas con características muy específicas, por ser muy atípicos los sujetos que participaron de la muestra; el valor de dicha investigación también se vería menoscabado.

Ahora cómo conseguimos que la muestra que vamos a utilizar sea representativa. Un punto importante para esto es de qué manera vamos a seleccionar a los sujetos que van a formar parte de la muestra. Hay distintos procedimientos para hacerlo, a los que vamos a denominar técnicas de muestreo. Según qué técnica de muestreo utilicemos, podremos tener mayor o menor seguridad, en cuanto a que la muestra sea representativa. Los procedimientos de muestreo los podemos dividir en dos grandes grupos, muestreos probabilísticos (aleatorios) y muestreos no probabilísticos (no aleatorios). Básicamente la diferencia entre ambos estaría dada por si se utiliza o no el azar para la elección de los sujetos que compondrán la muestra. De manera que *“en las muestras aleatorias cada elemento de población tiene una probabilidad igual, o una probabilidad cuantificable, de ser seleccionado”*, Clark-Carter, D. (2002) (pág. 159). En principio, los muestreos probabilísticos son los que nos garantizan de mejor manera que la muestra sea representativa. Además del elemento principal que es la elección al azar de los sujetos, hay otros procedimientos que se pueden emplear para garantizar de mejor manera la representatividad de la muestra, tal como menciona Pereda Marín, S. (1987); estos son: la estratificación y la proporcionalidad. De esta manera dentro de los muestreos probabilísticos tenemos, como menciona Cortada de Kohan, N. (1994): muestreo al azar simple, muestreo sistemático, muestreo estadificado al azar, muestreo estratificado proporcional, muestreo estratificado no proporcional y muestreo por conglomerados. En un muestreo al azar simple, utilizamos algún procedimiento al azar para elegir a los sujetos de la población que van a formar parte de la muestra. En el muestreo sistemático a partir de un primer caso elegido al azar, los restantes se eligen de un listado a intervalos sistemáticos. El muestreo por conglomerados se suele utilizar cuando la población tiene una distribución geográfica muy amplia, en ese caso en lugar de elegir al azar directamente a los individuos que componen la población, lo que se hace primero es elegir al azar grupos más amplios que se llaman conglomerados, como por ejemplo Municipios; y después se seleccionan al azar, de cada conglomerado, los individuos que van a conformar la muestra. En el caso del muestreo estratificado lo que se hace es *“dividir a la población en subpoblaciones o estratos y se selecciona una muestra para cada estrato”*, Hernández Sampieri, R.; Fernández Collado, C. y Lucio, P.B. (1997) (pág. 217). Cuando de cada estrato se selecciona la misma cantidad de sujetos es un muestreo estratificado simple. Cuando *“la representación de los estratos es ponderada por su varianza en la variable bajo estudio; los estratos más homogéneos aportan menos casos y los estratos menos homogéneos más casos”*, León, O.G. y Montero, I. (2003) (pág. 112), estamos frente a un muestreo estratificado óptimo como lo llaman León, O.G. y Montero, I. (2003) o muestro estratificado no proporcional como lo llama Cortada de Kohan, N. (1994). Este muestreo se suele aplicar cuando la variabilidad dentro de los estratos es muy desigual. En el caso del muestreo estratificado óptimo,

tenemos que tener mucho cuidado al generalizar los resultados, ya que la proporción de los sujetos en la muestra no respeta las existentes en la población. De manera que, según el tipo de estudio, y en todos los casos en estudios descriptivos, para poder generalizar habría que hacer las correcciones correspondientes, para ajustar los valores obtenidos a las proporciones reales que presentan los estratos en la población.

Cuando en la muestra se mantienen “*las mismas proporciones de sujetos que existen en la población dentro de cada estrato*”, Pereda Marín, S. (1987) (pág. 127), estamos frente a un muestreo estratificado proporcional, que reúne las tres condiciones para asegurar en forma plena que la muestra sea representativa: selección al azar, estratificación y proporcionalidad.

Cuando hablamos de que una muestra es representativa, lo que estamos diciendo es que “*los estadísticos que se obtienen en la muestra corresponden, representan, sin error sistemático, a los parámetros de la población*”, Pereda Marín, S. (1987) (pág. 126). Aclaremos un poco esto. Las medidas que obtenemos a partir de una muestra, por ejemplo, si calculáramos una media aritmética de alguna variable, las denominamos estadísticos. Si tomáramos a toda la población, para calcular dicha medida, eso sería un parámetro. Como por lo general, por una imposibilidad práctica, no realizamos un censo de la población, que sería tomar medidas de todos los componentes de la misma, tratamos de aproximar al valor del parámetro por medio del estadístico obtenido a partir de la muestra. Por lo general va a haber una diferencia entre el valor que se obtendría en la población (el parámetro) y el valor obtenido en la muestra (el estadístico); si bien en algún caso, podría ser que ambos coincidieran exactamente, aunque esto sería altamente improbable. Pero como mencionamos lo más usual es que esto no ocurra y que haya una diferencia entre ambos. Cuanto más cercanos estén el parámetro y el estadístico más representativa será la muestra. Obviamente para que esto ocurra es importante que en la muestra encontremos sujetos con características diferentes, tal cual se hallan en la población, de ahí que el azar, al asegurarme que cada sujeto de la población puede ser elegido, nos da una mayor seguridad de que se cumpla lo anteriormente mencionado. Además si le sumamos estratificación y proporcionalidad, a partir de variables previamente conocidas, que sean importantes para describir a la población de interés, podremos tener una mayor confianza en que la muestra sea adecuadamente representativa.

Cabe aclarar que un muestreo en el que haya estratificación y proporcionalidad, además de selección al azar, nos daría mayor seguridad de que la muestra sea representativa, pero de todos modos, cualquier tipo de muestreo probabilístico nos permite tener confianza sobre la representatividad de la muestra. Además, hay que tener en cuenta que “*la estratificación requiere un mayor esfuerzo y un mayor tiempo que la simple aleatorización*”, Pereda Marín, S. (1987) (pág. 130).

El estadístico es una estimación del valor del parámetro y “*como con cualquier estimado, nunca podemos estar totalmente seguros de que nuestro estimado de un parámetro es exacto. Sin embargo, si podemos encontrar un rango de valores que nos dé cierto nivel de confianza de que el parámetro debe caer en él. Tal rango se llama intervalo de confianza*”, Clark-Carter, D. (2002) (pág. 163).

Hasta acá hemos hablado de la importancia de que el muestreo sea probabilístico para tener una mayor seguridad de la representatividad de la muestra. Pero por otro lado tenemos que muchas veces se hacen muestreos no probabilísticos. Estos muestreos son más proclives a generar muestras sesgadas, o sea, no representativas. De manera que entre un muestreo probabilístico bien realizado y un muestreo no probabilístico, es mucho más probable que sea menos sesgada (más representativa) la muestra que utilizó el azar para la elección de los sujetos. Pero puede ocurrir, aunque sea más improbable, que una muestra no probabilística, no presente mucho sesgo y en consecuencia sea representativa.

En el caso de los muestreos probabilísticos hay que estar también al tanto de la cantidad de sujetos seleccionados que deciden no participar de la investigación. Esto es esperable, ya que uno selecciona los sujetos, pero después depende de su buena voluntad para que participen de la investigación, y muchos optan por no hacerlo. Supongamos que estamos haciendo una encuesta, ya sea telefónicamente, por correo, o entrevistando personalmente a las personas. Al solicitarles que participen de la encuesta, muchos de ellos, no querrán hacerlo. Y como dijimos, esto es esperable, y en principio, no tendría, porque afectar la representatividad de la muestra; siempre y cuando ese porcentaje de deserción sea bajo. Pero si es muy alto tendríamos que replantearnos si la muestra es realmente representativa, ya que no podríamos saber, si la pérdida de sujetos, se dio en forma aleatoria, lo cual no afectaría la representatividad de la muestra, o si por el contrario, respondió a algún o algunos factores sistemáticos. Dando por resultado, que los sujetos que terminaron participando de la muestra tengan en alguna o algunas variables, características diferenciadas de los que no participaron y que justamente estas características diferenciadas fueron las que dieron lugar a la no participación de algunos de los sujetos. Por ese motivo es importante, en aquellos casos en que se pueda hacer (por ejemplo si el contacto es personal), identificar y registrar que características tienen los sujetos que no quieren participar, por ejemplo: sexo, edad, etc. Esto permitiría, aunque sea mínimamente, analizar si en esas características tales sujetos se diferencian o no del conjunto de los integrantes de la muestra. Y hacer entonces, de ser necesario, las correcciones pertinentes o extraer conclusiones acerca de cómo se pudo haber visto afectada la representatividad de la muestra.

En cuanto a encuestas tomadas utilizando muestras no probabilísticas o aleatorias, según como se realicen, se puede tener alguna idea sobre que presenten un mayor o menor sesgo. Si al sujeto que participa se le solicita que lo haga, de manera que en primera instancia es pasivo en cuanto a la voluntad de participar de la encuesta, simplemente acepta hacerlo o no ante el pedido del entrevistador (por ejemplo que el entrevistador se ubique en algún lugar y les solicite a quienes entren o pasen por allí si quieren contestar la encuesta), seguramente tendrá menor sesgo que si el sujeto decide activamente participar de la misma; por ejemplo alguna encuesta por Internet o telefónica donde es el sujeto el que tiene que llamar a algún número telefónico (ya que si al sujeto lo llaman para hacerle la encuesta, ya sería como el primer caso, donde este es pasivo, en cuanto a la iniciativa de participar). ¿Por qué motivo se daría esta diferencia en el mayor o menor sesgo de una y otras encuestas? Podemos decir que una cosa es responder afirmativamente a una solicitud que alguien nos hace, y otra muy diferente es tener la motivación de participar por propia iniciativa, lo que ocurriría en los

casos mencionados de Internet o que el sujeto llame por teléfono. En esa motivación siempre se encuentran intereses de la persona respecto de ese tema, experiencias positivas o negativas que lo mueven a participar y que lo diferencian del resto de la población que no lo hace.

Del mismo modo, seguramente presentará menor sesgo una encuesta en la cual se diversifican las ubicaciones y los horarios donde se entrevista a los sujetos, o una encuesta idéntica que se hace en distintas páginas de Internet (que tengan temáticas diferentes), que si se recoge información en un solo lugar y horario, o en una sola página de Internet, programa de radio o televisión. En estos últimos casos, seguramente, la información que se recoge se limita a la población que suele concurrir a ese lugar y en ese horario, o que lee esa página de Internet, o escucha ese programa de radio, lee esa revista o mira ese programa de televisión, en los cuales se realiza la encuesta; pero es muy improbable que sea representativa de la población general. Por ejemplo, supongamos que quiero hacer una investigación sobre estudiantes universitarios y realizo un muestreo accidental ubicándome en la entrada de la Facultad de Psicología de la UBA, pidiéndoles a los alumnos que ingresan al edificio que contesten una encuesta. Si mi población de interés fueran los estudiantes de psicología de la UBA, estaría bien, siempre y cuando se hubiera tenido el cuidado de cubrir los distintos horarios, ya que el perfil del alumnado puede variar según la banda horaria a la que asistan. Pero si mi población de interés fueran los estudiantes de psicología en general, tendría que diversificar los lugares en los cuales recogiera información, tanto más cuanto más extensa geográficamente o cualitativamente sea la población de interés. Por ejemplo: estudiantes de psicología de una ciudad, de una provincia o de todo el país; de universidades públicas, privadas o ambas. Según la opción que tome más amplitud, tendré mi población, y en consecuencia habrá que hacer una mayor diversificación de los lugares de selección de los sujetos de la muestra.

Otra forma de mejorar un muestreo no probabilístico es hacer una muestra de cuota, en la cual hay *“cuotas establecidas para el número de personas que habrán de incluirse en las submuestras. Por ejemplo, tal vez los investigadores deseen un número igual de hombres y mujeres. Una vez que obtengan su cuota para un género, se acercarán solo a miembros del otro género, hasta que tengan los individuos suficientes”*, Clark-Carter, D. (2002) (pág. 162). Se pueden establecer cuotas en función de varias variables que interesen a los investigadores, lo cual mejora aún más el procedimiento de muestreo.

Además, si en el informe posterior donde se describe el trabajo realizado, se detalla todo el procedimiento de muestreo realizado, quien lo lee tendrá una idea acabada de en cuanto y a quienes se puede generalizar esa investigación.

Hay un caso particular de muestreo no probabilístico que es el muestreo a propósito. *“En ocasiones son expertos los que hacen una selección a propósito cumpliendo el objetivo de que los integrantes de la muestra cumplan algún requisito. Suponga que estamos en la fase inicial de construcción de un test que va a medir el grado de desarrollo de la escritura en niños. Necesitamos una muestra de los distintos niveles de desarrollo en el grupo de edad para el que vamos a construir el test. Una forma de rastrear la variabilidad*

de los niños es pedir a varios maestros, de distinto tipo de colegio y distinto tipo de entorno, que nos proporcionen los dos mejores niños, los dos peores y dos que estén en la media”, León, O.G. y Montero, I. (2003) (pág. 116). En este caso lo que interesa no es tanto elegir los sujetos al azar para que la muestra sea representativa, sino que los mismos cumplan con ciertas características en función de necesidades de la investigación.

Además, en investigación a veces se trabaja con muestras incidentales. “*Se denomina muestra incidental a aquella que se emplea porque está a disposición del investigador en un momento determinado*”, Pereda Marín, S. (1987) (pág. 127).

Vamos a pasar a enfocar ahora la cuestión de otra manera. Hasta acá hemos venido hablando de lo que sería ideal desde un punto de vista metodológico. Pero la realidad práctica de una investigación a veces va por otros caminos. Entonces nos abocaremos a analizar, cuánto de lo que refiere a una muestra representativa se cumple en los distintos métodos de investigación cuantitativos, y en qué medida ciertas libertades en cuanto al muestreo pueden repercutir o no en la calidad de la investigación y en particular en su validez externa.

En general, en investigación cuantitativa podemos hablar de los siguientes métodos de investigación: descriptivo, correlacional y explicativo (experimento y cuasiexperimento). No incluimos los diseños preexperimentales, los cuales a los fines de este trabajo podemos tratarlos como cuasiexperimentos con menor validez interna; y como ya comentáramos con anterioridad, tampoco incluimos a los diseños de sujeto único. Aclaramos que todos estos métodos responden a una misma lógica metodológica, que son complementarios entre sí, y que en última instancia, las diferencias estriban en cuestiones de diseño, validez y objetivos de la investigación.

Cuando estamos frente a un método descriptivo, el único propósito es describir un fenómeno en términos cuantitativos, y aunque se utilice más de una variable no se establece ningún nexo entre ellas. Por ejemplo si realizáramos un estudio epidemiológico sobre la prevalencia de psicopatologías en determinada región, pero sin vincular estas con ninguna otra variable. Cuando un estudio es descriptivo es fundamental que la técnica de muestreo empleada nos dé seguridad de que la muestra sea representativa. Ya que si el único propósito de la investigación es describir y la muestra estuviera sesgada, de tal manera que los valores obtenidos de la misma no representaran a la población a la cual refieren, se informaría mal lo único que la investigación pretende transmitir.

De todos modos, consideramos que en algunos casos donde no haya ninguna o muy poca información descriptiva sobre un tema en particular; y considerando que cuanto más complicada la técnica de muestreo requiere más esfuerzo, tiempo y tiene mayor costo; lo cual hace que a veces no se disponga de lo medios necesarios para un muestreo probabilístico. Una muestra no aleatoria es preferible a un vacío de información. Siempre y cuando esa muestra no aleatoria sea realizada de la mejor forma posible, y se informe adecuadamente sobre los procedimientos empleados, para que quienes lean los resultados puedan extraer sus propias conclusiones. También sería importante si se dispone de datos sobre el tema en estudio, pero de una investigación realizada en otro ámbito geográfico, comparar los resultados de ambas investigaciones.

Ahora, en los restantes tipos de investigaciones: correlacionales y explicativas, y especialmente en estas últimas, el hecho concreto es que mayormente las muestras que se toman no son probabilísticas. Como dice Pereda Marín, S. (1987) (pág. 127): *“en Psicología, la mayoría de las investigaciones no se llevan a cabo con muestras representativas, sino con muestras incidentales”*. Además de eso, como menciona el mismo autor, hay un uso generalizado de estudiantes de psicología como sujetos de las investigaciones: *“la mayoría de las investigaciones que se llevan a cabo en Psicología emplean como sujetos a estudiantes de esta carrera, lo que constituye un ejemplo masivo de muestras incidentales y da lugar a la no posibilidad de generalización de los resultados obtenidos en dichas investigaciones”*, Pereda Marín, S. (1987) (pág. 128).

Vamos a analizar cómo se da lo anterior en cada caso y cómo podría afectar la validez externa de los resultados, en cuanto a la generalización de los resultados a la población. Lo anteriormente expuesto se puede aplicar tal cual a las investigaciones experimentales. Y este es el primer caso que vamos a analizar. Si bien el ideal metodológico sería tomar una muestra al azar, en un experimento uno se ve forzado a trabajar con voluntarios, la persona tiene que tener una voluntad de participación, y esto hace que en la práctica resulte casi imposible tomar una muestra aleatoria. Imaginemos la situación, elegimos al azar a determinada cantidad de sujetos de la población, ¿cómo hacemos para que acepten participar de la investigación?; ya que no se trata de contestar una encuesta, sino de participar de determinados procedimientos experimentales, algunas veces bastante complejos. La inmensa mayoría de los sujetos seleccionados, simplemente nos dirían que no desean participar de la investigación; y con semejante pérdida de sujetos, la supuesta muestra representativa, no lo sería en lo más mínimo. Además de la enorme dificultad que tendríamos para poder formar una muestra con una aceptable cantidad de sujetos.

Lo que sí se hace en un experimento y, justamente, es un requisito indispensable para que una investigación sea considerada un experimento, es la asignación al azar de los sujetos a los grupos. Pero esto es completamente diferente a tomar una muestra al azar, ya que simplemente se asigna mediante un procedimiento aleatorio, cada sujeto que ya forma parte de la muestra, a alguno de los grupos existentes en el diseño de la investigación, independientemente de la forma en que se haya constituido la muestra

La pregunta es: ¿Qué los sujetos de la muestra no hayan sido seleccionados al azar afecta realmente la capacidad de generalizar los resultados?, y de ser así, ¿en cuánto compromete la validez externa de la investigación?

En primer lugar, tenemos que tener en cuenta que en las investigaciones experimentales estamos relacionando variables y explicando teóricamente, no solo describiendo. Estas explicaciones de por sí ya plantean cierto grado de generalidad; además no toda variable de sujeto será relevante al fenómeno en estudio, y aquellas que sí fueran relevantes, deberían haber sido controladas adecuadamente, para preservar la validez interna de la investigación. Si fueron controladas adecuadamente, los resultados obtenidos y las generalizaciones que se hagan de los mismos ya las contemplan. Si no fueron controladas

adecuadamente, el problema no solo tendría que ver con la validez externa del experimento, sino con algo que es en tales investigaciones sumamente importante, su validez interna. De todos modos, para no volver a confundir lo ideal, con la realidad de la práctica, el control de las variables contaminadoras es una cuestión de grado, y en general, se está lejos de llegar al control absoluto de todas las variables potencialmente relevantes.

Debemos considerar además que en el informe de la investigación se refiere con precisión en el apartado Sujetos, que características tenía la muestra en cuanto a género, edad, nivel socioeconómico, nivel de estudios alcanzado, etc. Esto nos da una clara idea de a qué población puede generalizarse la investigación. Por ejemplo si la muestra estuvo compuesta exclusivamente por mujeres, solo podremos generalizar a ese grupo poblacional. También en esto, debemos tener en cuenta el tema en estudio y cuánto se supone teóricamente que este puede verse afectado por variables sociodemográficas. Si concluyéramos que puede verse afectado por las mismas, deberíamos ser muy cuidadosos en cuanto a quienes generalizamos. Si es muy improbable que dichas variables sociodemográficas afecten al fenómeno en estudio, el problema de generalización que se plantearía sería mucho más limitado.

En ciencia es importante el tema de poder replicar las investigaciones. Indirectamente, esto también colabora, en parte, a solucionar este problema de muestras no probabilísticas, aunque la replicación no persigue básicamente ese propósito, sino el de probar si las hipótesis planteadas, pueden corroborarse. Al tomar distintas muestras, para realizar distintas replicaciones, se aumenta la variedad de los sujetos que componen las mismas, con la consiguiente probabilidad, de que en el conjunto de esas investigaciones se aproxime más a las características de la población.

De tal manera podemos decir que, en general, el tema de que las muestras no sean probabilísticas, no afectaría tanto la validez externa de un experimento. En realidad el problema de validez externa de los experimentos, no tiene que ver únicamente con la representatividad de las muestras, sino también, y fundamentalmente con su validez ecológica.

Un comentario aparte merece la cuestión de que los integrantes de la muestra sean exclusivamente estudiantes de psicología. En este caso consideramos que deberían manejarse los resultados obtenidos con precaución, en cuanto a su generalización, hasta tanto se pudieran replicar con muestras que incorporen sujetos que no sean estudiantes de psicología; o que se analice, si la categoría estudiante de psicología forma parte de una variable relevante para el fenómeno en estudio, lo cual para hacerlo, presupone tomar muestras que no estén formadas exclusivamente por estudiantes de psicología. Esto habría que tenerlo en cuenta principalmente si la o las variables en estudio, pueden verse afectadas por variables sociodemográficas, o por variables de sujeto que puedan incidir en la elección de la carrera universitaria. De no ser así, no sería tan problemático que la muestra estuviera constituida exclusivamente por estudiantes de psicología, lo cual no quita que replicar la investigación con otros sujetos, seguiría siendo sumamente útil.

Lo anteriormente mencionado para los experimentos podemos extenderlo para los cuasiexperimentos con manipulación artificial de la variable independiente.

En el caso de los cuasiexperimentos con manipulación natural de la variable independiente y en las investigaciones correlacionales, podemos decir que también estamos relacionando variables y no solo describiendo; y específicamente en el cuasiexperimento, además, explicando teóricamente. También se describe con precisión el perfil de la muestra. Dichas investigaciones también es deseable que sean replicadas. De manera que en estos casos también podemos aplicar con las adaptaciones a cada método lo dicho para los trabajos experimentales. Cabe aclarar que una diferencia importante que tienen las investigaciones correlacionales, respecto a los experimentos, en cuanto a su validez externa, estriba en que en las investigaciones correlacionales se trabaja exclusivamente con operacionalizaciones de medida. Una *“definición operacional de medida describe cómo se medirá una variable. Indicará, por tanto, qué instrumentos se emplearán en la medida, en que unidades se hará, etc.”*, Pereda Marín, S. (1987) (pág. 165). De tal manera que, como en estas investigaciones (correlacionales), en el caso de todas las variables se realiza tan solo una recogida de información para medirlas, pero en ningún caso se las manipula artificialmente, el contexto del estudio es más natural y más directamente trasladable a situaciones naturales, lo que hace que tengan una mejor validez externa ecológica. Lo mismo es aplicable a los cuasiexperimentos con manipulación natural, ya que al no ser la variable independiente manipulable en forma artificial, solo se realiza de la misma una operacionalización de medida, tendiente a medir la variable para posteriormente clasificar a los sujetos en función de las categorías de la variable independiente. En el caso de la variable dependiente siempre tiene una operacionalización de medida.

Podemos concluir entonces, que el hecho de que las muestras no sean probabilísticas, tampoco afectaría tanto la validez externa de las investigaciones anteriormente mencionadas; y que de hecho tendrían más validez externa que un experimento, gracias a la buena validez ecológica que tienen.

Tamaño de la muestra

Hay una relación entre el tamaño de la muestra y la medida en que la misma es representativa. *“Tener una muestra que contenga las características de la población no es suficiente. Necesita tener un cierto tamaño para que quede libre de esos errores que pueden ocurrir por azar y anularían la representación de la muestra”*, León, O.G. y Montero, I. (2003) (pág. 110). *“Ciertamente, la única muestra igual a la población es la población misma. A medida que se aumenta el tamaño de la muestra, se irán incluyendo más y más sujetos con diferentes aspectos que caracterizan a la población”* Pereda Marín, S. (1987) (pág. 126). Además el tamaño de la muestra tiene una directa relación con el intervalo de confianza que se plantea para las estimaciones de la muestra. *“Cuanto mayor sea la muestra, tanto menor será el rango del intervalo de confianza para el mismo nivel de confianza; es decir, podemos determinar con mayor exactitud el parámetro de la población”*, Clark-Carter, D. (2002) (pág. 165).

Para que una muestra sea representativa no bastaría entonces con que los sujetos de la misma sean elegidos al azar, se necesitaría también contar con determinada cantidad de sujetos que integren la muestra, especialmente si queremos reducir los márgenes de error de la misma. Para determinar en cada caso el tamaño de muestra requerido se utilizan procedimientos estadísticos, cuya exposición no es objeto de este artículo.

Esto no quiere decir que siempre lo deseable sea, necesariamente, aumentar el tamaño de la muestra para reducir el rango del intervalo de confianza, y obtener medidas más precisas al tener un menor margen de error. En un sentido práctico se podría decir que si se tiene una muestra con pocos sujetos, sí sería deseable incrementar el tamaño de la misma, ya que eso reduciría considerablemente el margen de error de las estimaciones. Ahora, si el tamaño de la muestra ya es grande, incrementar el tamaño de la misma, solo conseguiría una pequeña reducción en el margen de error, y para reducir en más el margen de error se requeriría un aumento realmente enorme en el tamaño de la muestra. Esto es así porque *“cuanto mayor sea la muestra, tanto mayor será el aumento del tamaño de la muestra que se precisa para reducir el intervalo de confianza en una cantidad equivalente”*, Clark-Carter, D. (2002) (pág. 165).

De manera que si tuviéramos una muestra de 200 sujetos, con determinado intervalo de confianza y el consiguiente margen de error (el margen de error es la mitad del intervalo de confianza), y agregamos 300 sujetos más a la muestra, pasando a tener 500 sujetos, conseguiríamos una importante reducción en el intervalo de confianza. Ahora si volviéramos a aumentar la muestra en 300 sujetos, pasando a tener una muestra de 800 sujetos, la reducción que obtendríamos en el intervalo de confianza sería mucho menor. Si consideramos el esfuerzo que requiere conseguir una muestra grande; si uno ya dispone de un tamaño de muestra importante, y técnicamente apropiado, habría que plantearse si vale la pena realizar un gran esfuerzo adicional para incrementar el tamaño de la misma, si solo conseguiríamos con ello una pequeña reducción en el margen de error. Pero, como veremos más adelante, el tamaño de la muestra tiene otras ventajas aparte de la reducción en el margen de error de las estimaciones.

Pero hay que tener en cuenta, para diferenciar nuevamente entre lo ideal y la práctica, que de hecho en psicología en muchas investigaciones se trabaja con muestras pequeñas, sin mencionar el caso de los diseños de sujeto único, que quedan por fuera de este artículo. En lo que respecta a lo que esto afectaría a la representatividad de la muestra, consideramos que podemos aplicar los mismos criterios que habíamos planteado previamente para el caso de los muestreos aleatorios o no aleatorios. Si la investigación es descriptiva el tamaño de muestra es sumamente importante, si la investigación fuere correlacional o explicativa, un tamaño de muestra importante sería deseable, pero no condición excluyente para la calidad de la investigación en lo que hace a su validez externa.

Pero tenemos que tener en cuenta que el tamaño de la muestra no es solo importante por el tema de la representatividad sino que hay otras cuestiones metodológicas y estadísticas por las cuales un mayor tamaño de muestra puede ser un requisito importante para una investigación y esto sí se aplicaría en forma exclusiva para las investigaciones correlacionales y explicativas, y no sería relevante para las investigaciones descriptivas.

Tamaño de muestra y diseño de investigación

Según las características del diseño de investigación que tengamos, vamos a precisar tamaños de muestra mínimos para que los mismos puedan llevarse a cabo. Fundamentalmente por cuestiones estadísticas, ya que una vez implementados dichos diseños tiene que llevarse a cabo el análisis estadístico correspondiente y aquí nos encontraríamos en algunos casos con limitaciones técnicas, si no contáramos con el tamaño de muestra apropiado. Lo anterior se aplica fundamentalmente a diseños multivariados. Por ejemplo:

- Diseños que tienen más de una variable independiente;
- También en el caso de trabajos correlacionales, donde se trabaje con conjuntos de predictores;
- Y donde se apliquen cálculos tendientes a conseguir una reducción de datos, como por ejemplo, el análisis factorial.

Tomaremos para ejemplificar los siguientes casos:

Por un lado, investigaciones que utilicen para el cálculo estadístico regresión lineal múltiple, o que utilicen, regresión logística. En ambos casos para la confiabilidad del cálculo necesitamos tener determinada cantidad de sujetos en función de la cantidad de predictores incluidos en el análisis. Hay distintos criterios en cuanto a la cantidad de sujetos necesarios para la regresión lineal múltiple. Como menciona Gardner, R.C. (2003), los mismos van desde indicar una cantidad mínima de 5 sujetos por predictor, hasta sugerencias de 30 sujetos por predictor, mientras que algunos sugieren que siempre hay que tener por lo menos una muestra de 400 sujetos. Consideramos que un número mínimo conveniente podrían ser diez sujetos por variable involucrada en el análisis, tanto para la regresión lineal múltiple, como para la regresión logística. Hacer el cálculo con una muestra menor no sería aconsejable. Disponer de una muestra aun mayor, a los fines estadísticos, siempre es mucho mejor, dado que los resultados se vuelven más estables. Hay que considerar que en este caso el tamaño de la muestra es relativo a la cantidad de predictores. Si tomamos dos muestras de igual tamaño, pero que pertenecen a investigaciones con distinta cantidad de predictores, la que tenga menos predictores tendrá un tamaño de muestra relativamente más importante.

Para la construcción de instrumentos psicométricos se utiliza mucho el análisis factorial, que es una técnica de reducción de datos. *“El principal propósito del análisis factorial es reducir un conjunto amplio de variables a un reducido grupo de medidas más fácil de manipular. El análisis factorial desentraña las complejas interrelaciones existentes entre variables e identifica cuáles de estas se enlazan en torno de conceptos unificados. Las dimensiones fundamentales así identificadas se denominan factores”*, Polit, D.F. y Hungler, B.P. (1997) (pág. 448).

“En general las estructuras factoriales pueden tender a ser inestables y dependen mucho del tamaño de la muestra. El tamaño necesario depende en parte de la cantidad de ítems involucrados en el análisis, y hay diferentes criterios para establecer el número de sujetos indicado. Uno de los más extendidos, es que el número de sujetos no

sea menor a cinco veces la cantidad de ítems del instrumento. Aun con este número de sujetos, las estructuras factoriales pueden seguir siendo inestables", Argibay, J.C. (2006) (pág. 29). Justamente por este motivo es que hay criterios que indican todavía mayor cantidad de sujetos por variable, cinco sería el número mínimo requerido. Aún con este número mínimo, si tuviéramos que hacer el análisis factorial de un *test* que tuviera 100 ítems, requeriríamos mínimo 500 sujetos.

Por último vamos a referirnos al tamaño de la muestra en diseños factoriales. Un diseño factorial es una investigación en la cual tenemos más de una variable independiente. En un diseño factorial se analizan tanto los efectos aislados que producen las variables independientes, como los efectos combinados que producen las mismas. Las variables independientes involucradas en el diseño factorial pueden ser tanto de manipulación artificial, como de manipulación natural. En aquellos casos que alguna de las variables independientes tuviera manipulación natural, estaríamos frente a una investigación cuasiexperimental, caso contrario sería una investigación experimental.

En un diseño factorial con medidas independientes, *"aquellos en que los efectos de cada tratamiento experimental son probados en un grupo diferente de sujetos"*, Pereda Marín, S. (1987) (pág. 355); vamos a tener tantos grupos como combinaciones de los niveles de las variables independientes. Hay que tener en cuenta que para la correcta aplicación del análisis estadístico correspondiente, sería recomendable tener como mínimo diez sujetos por grupo. Esto muchas veces establece una limitación en cuanto a la cantidad de niveles a utilizar por factor (variable independiente). Ya que si tomamos muchos valores para cada uno de ellos, el número de combinaciones aumenta mucho, con la consiguiente necesidad de tener mayor cantidad de grupos, y la necesidad entonces de disponer de una muestra más grande. Por ejemplo: si tuviéramos un diseño factorial con tres factores de medidas independientes, y para cada uno de ellos quisiéramos tomar tres niveles, tendríamos 27 grupos resultantes de la combinación de los factores y precisaríamos 270 sujetos mínimo, para poder implementar adecuadamente el diseño. Si no dispusiéramos de tal tamaño de muestra, tendríamos que adjudicar a los factores menos valores. Esto es aplicable directamente como dijimos a los diseños factoriales con medidas independientes en todos los factores, ya que si el diseño factorial fuera de medidas repetidas en todos o en algunos factores, el tema de la cantidad de grupos resultante de la combinación de los factores sería diferente.

El tamaño de la muestra, si es pequeño, también impondría limitaciones a la cantidad de factores que se podrían incluir en el diseño. Aunque vale aclarar en este caso que por características del diseño mismo no se suelen usar diseños factoriales que tengan más de cuatro factores.

Si el diseño factorial fuera cuasiexperimental por ser todos los factores o alguno de ellos de manipulación natural, el tema se complica aún más; principalmente cuando todos los factores o más de uno de ellos solo permitieran un nivel de manipulación natural. Cuando la manipulación es artificial, uno decide qué sujetos de la muestra va a asignar a cada grupo, y de esa manera suelen quedar grupos con igual cantidad de sujetos. Pero cuando la manipulación es natural, tomamos una muestra y luego clasificamos a los sujetos según los valores que presentan, o sea, que la asignación a cada

grupo viene dada naturalmente y no por decisión del investigador, lo cual hace que los grupos, por lo general, queden integrados por un número desigual de sujetos. De manera que si decidimos el tamaño de muestra que vamos a necesitar mínimamente, de la manera anteriormente expuesta, cantidad de grupos por diez, seguramente ocurriría que tendríamos varios grupos con sujetos de más y otros grupos con menor cantidad de sujetos que los requeridos, ya que esto sería resultado de cómo se den naturalmente en cada sujeto la/s característica/s que usemos como variables independientes en la investigación. Esto hace que en estos casos, se tenga que tomar una muestra bastante mayor, que la establecida según la cantidad de grupos, para disponer en cada grupo, como mínimo, del número de sujetos requerido.

Tamaño de muestra y significación estadística

Al analizar los resultados de una investigación tomamos decisiones sobre si corroboramos o no, la o las hipótesis planteadas en la investigación, y en investigaciones exploratorias, simplemente si las variables en estudio están relacionadas. Para ello tenemos que recurrir a una decisión estadística, que implica la prueba de la hipótesis estadística. En este modelo de decisión estadística, un paso importante es comparar el valor de significación obtenido por medio de los análisis estadísticos correspondientes, con el nivel de significación o probabilidad crítica (alfa), establecido previamente. De manera que dependemos para poder corroborar nuestras hipótesis o relaciones entre variables, de que la probabilidad de que nuestro resultado haya ocurrido por azar (significación), obtenida en los cálculos estadísticos realizados, sea igual o menor que alfa. De manera que el valor de significación obtenido en los cálculos estadísticos, es de fundamental importancia para poder corroborar lo planteado en las investigaciones que realicemos. Pero ocurre que la significación que obtengamos esta muy condicionada por el tamaño de muestra que empleemos. De manera que *“dos estudios conducidos de la misma manera en todos los aspectos, excepto el tamaño de la muestra, pueden llevar a conclusiones diferentes. El que tiene la mayor muestra puede alcanzar un resultado estadísticamente significativo, mientras que el otro no... De acuerdo con esto, cuanto más pequeño sea el tamaño de la muestra, tanto más probable resultará que cometamos un error tipo II (rechazando la hipótesis de investigación cuando en realidad es correcta)”*, Clark-Carter, D. (2002) (pág. 186).

De manera que el hecho de que podamos corroborar nuestras hipótesis, además de que sean correctas, depende bastante del tamaño de muestra que empleemos. Obviamente, esto esta condicionado también por el tamaño del efecto, o sea, si la variable independiente que estamos estudiando produce un efecto considerable o un efecto pequeño sobre la variable dependiente, o si, la relación entre las variables que estudiamos es grande, mediana o pequeña. Esto podríamos traducirlo a porcentaje, diciendo cuánto es el porcentaje que una variable está explicando de la variabilidad de otra. Cuando el porcentaje de varianza explicada es alto, una muestra aunque sea de tamaño pequeño puede detectarlo; ahora, cuando el porcentaje de varianza explicada es bajo, se requerirán muestras grandes para poder alcanzar valores de significación estadística que nos permitan corroborar las hipótesis planteadas.

Podemos decir, entonces, que los valores de significación obtenidos en dos investigaciones diferentes, con tamaños de muestra diferentes, no son comparables entre sí,

para determinar cuál investigación tiene un resultado de mayor magnitud. Para ello debemos usar las medidas de tamaño del efecto que son independientes del tamaño de la muestra.

De manera que el valor de significación que podamos obtener esta muy condicionado por el tamaño de la muestra. Muestras más grandes hacen más difícil que podamos cometer un error de tipo II, al extraer conclusiones sobre una investigación a partir del análisis de datos. Por eso, se puede analizar el poder de la prueba estadística para decidir, durante la etapa de diseño de la investigación, el tamaño de muestra requerido para reducir la probabilidad de cometer un error de tipo II.

También hay que tener en cuenta que muestras muy grandes, a veces, nos permiten detectar efectos que, si bien son válidos, pueden implicar magnitudes de efecto muy pequeñas. Es importante, entonces, al hacer el análisis de datos, calcular además de la significación (que nos permite decidir sobre la hipótesis planteada), el tamaño del efecto, que nos permite saber cuánto explican una o varias variables de la variabilidad de otra.

Tamaño de muestra y formación de los grupos por aleatorización

En las investigaciones experimentales, un requisito básico, cuando se trabaja con diseños que incluyen variables con medidas independientes, es que los grupos que haya que formar para la aplicación del diseño de investigación, sean lo mas homogéneos posible, previo a los tratamientos (aplicación de los valores de la variable independiente o de las combinaciones de los niveles de varias variables independientes). Esto se hace para que las diferencias que uno encuentre con posterioridad puedan ser atribuidas a los tratamientos y no al hecho de que ya existieran previamente. Para poder tener esa seguridad sobre la equivalencia inicial de los grupos, se utiliza la aleatorización, o sea, asignar los sujetos a los grupos utilizando algún procedimiento al azar. Ahora para que esto ocurra adecuadamente, de manera que el azar pueda servir para repartir en forma equivalente a los sujetos en los grupos, dependemos también del tamaño de la muestra, ya que cuando son pocos casos, uno puede esperar una mayor variabilidad, de manera que el azar no nos daría tanta seguridad de que los sujetos fueron repartidos homogéneamente en los grupos. Para que el uso del azar nos de una mayor garantía de que los grupos son inicialmente homogéneos, se requieren muestras grandes. De disponer únicamente de pocos sujetos, es más conveniente no confiar tanto en la equivalencia inicial de los grupos, y emplear además de la aleatorización, medidas pretratamiento, o un diseño de bloques o un diseño de grupos apareados.

Conclusiones

Como hemos visto, el tema de la “muestra” es importante por distintos motivos. Por un lado, y que es quizá la primera cuestión que se piensa cuando se trata con muestras, que los valores que obtengamos de esa muestra nos permitan conocer los valores que encontraríamos en determinada población. Cuestiones como la forma en que se seleccionen los sujetos y el tamaño de la muestra, tienen una importancia fundamental al momento de poder determinar en qué medida esa muestra es representativa de la población a la cual refiere. Que el muestreo sea probabilístico, es un tema sumamente importante en el proceso de garantizar una muestra no sesgada, lo cual, en principio, se vincula con la validez externa de la investigación.

Consideramos importante diferenciar entre lo que sería ideal desde un punto de vista técnico metodológico, y aquello que ocurre de hecho en la tarea de investigar. Mencionamos, que suele ocurrir, que el muestreo no sea probabilístico en las investigaciones correlacionales y principalmente en las explicativas (experimento y cuasiexperimento). Podemos considerar que en realidad esto no afectaría demasiado la validez externa de las mismas, siempre y cuando se tomen algunas precauciones, y a partir del perfil de la muestra y del tema de investigación se pueda determinar a qué población, que tenga cuales características, se puede generalizar la investigación. Tenemos que tener en cuenta que en este caso nos referimos exclusivamente al aspecto de la validez externa, que tiene que ver con determinar a qué sujetos se pueden generalizar los resultados de la investigación. No nos referimos a la validez externa, en su aspecto ecológico, que hace a la posibilidad de generalizar a situaciones naturales los resultados obtenidos.

Es importante considerar que, si bien el tema del tamaño de la muestra es sumamente relevante, al momento de determinar que esta sea representativa, no es menos cierto que la cuestión del tamaño de la muestra excede esta cuestión, e involucra cuestiones muy importantes que hacen a la realización de una investigación.

Hay que tener en cuenta el tamaño de muestra con el que se va a contar y cuál es el que se requiere para poder implementar adecuadamente determinados diseños de investigación. Además, en lo que hace a una adecuada equivalencia inicial de los grupos de una investigación, conseguida mediante asignación al azar de los sujetos a los grupos (que es fundamental en un trabajo experimental); el tamaño de la muestra, si es adecuadamente grande, nos permite tener una mayor garantía de que se haya conseguido que los grupos sean homogéneos.

Por último, si queremos poder corroborar nuestras hipótesis de investigación. Además de que sean correctas, dependemos para poder hacerlo del tamaño de la muestra, ya que la significación estadística, que es necesaria para poder corroborar o no dichas hipótesis, depende y en mucho de la cantidad de sujetos que se incluyan en el análisis de datos. Al tener muestras con un tamaño insuficiente, corremos el riesgo de cometer un error de tipo II, o sea, rechazar la hipótesis de investigación, siendo esta realmente cierta. Cuando la muestra es muy pequeña, corroborar la hipótesis de investigación se hace sumamente difícil, más allá de que esta sea correcta. Lo cual nos da una clara idea de la ventaja de trabajar con muestras grandes.

Bibliografía

Argibay, J.C. (2006), "Técnicas psicométricas. Cuestiones de validez y confiabilidad". En: *Subjetividad y Procesos Cognitivos*, 8, 15-33.

Arнау Gras, J. (1982), *Psicología experimental*, México, Trillas.

Clark-Carter, D. (2002), *Investigación cuantitativa en psicología. Del diseño experimental al reporte de investigación*, México, Oxford University Press.

Cortada de Kohan, N. (1994), *Diseño estadístico*, Buenos Aires, Editorial Universitaria de Buenos Aires.

Gardner, R.C. (2003), *Estadística para Psicología usando SPSS para Windows*, México, Pearson Educación de México.

Hernández Sampieri, R.; Fernández Collado, C. y Lucio, P.B. (1997), *Metodología de la Investigación*, Colombia, McGraw-Hill.

León, O.G. y Montero, I. (2003), *Métodos de investigación en psicología y educación*. Madrid, McGraw-Hill/Interamericana de España.

Pereda Marin, S. (1987), *Psicología experimental. I. Metodología*, Madrid, Ediciones Pirámide.

Polit, D.F. y Hungler, B.P. (1997), *Investigación científica. En ciencias de la salud*, México, McGraw-Hill Interamericana.

Fecha de recepción: 19/02/09

Fecha de aceptación: 02/04/09